# Atomistic and reduced models of protein folding and aggregation

Development of coarse-grained models for protein simulations

Protein fold recognition - the protein structure prediction problem

Potentials for off-lattice Monte Carlo simulations of proteins

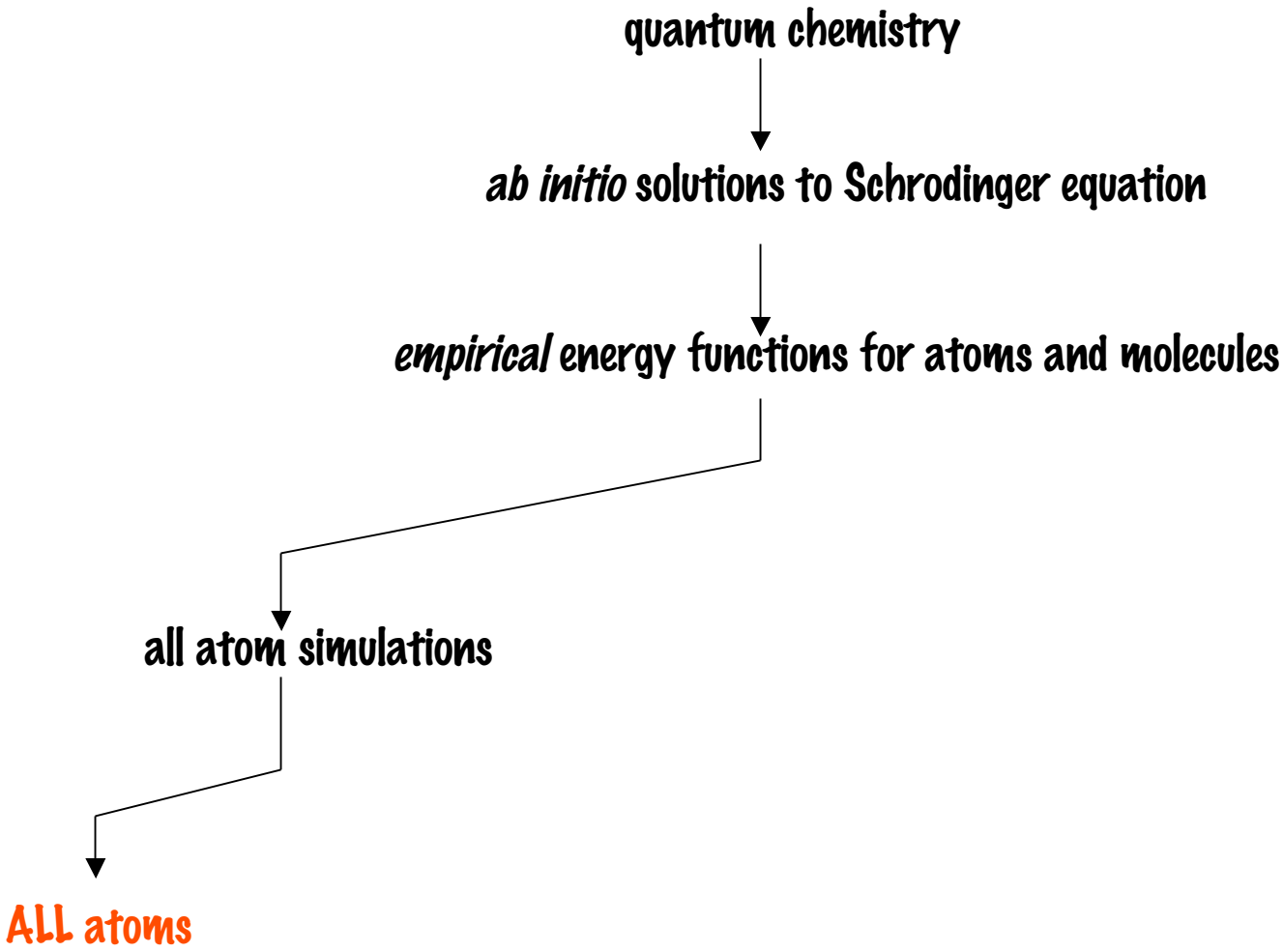
Application to a coil-to-helix transition in polyalanine

Vio Buchete, Alan van Giessen and Dave Thirumalai

NIH and NSF

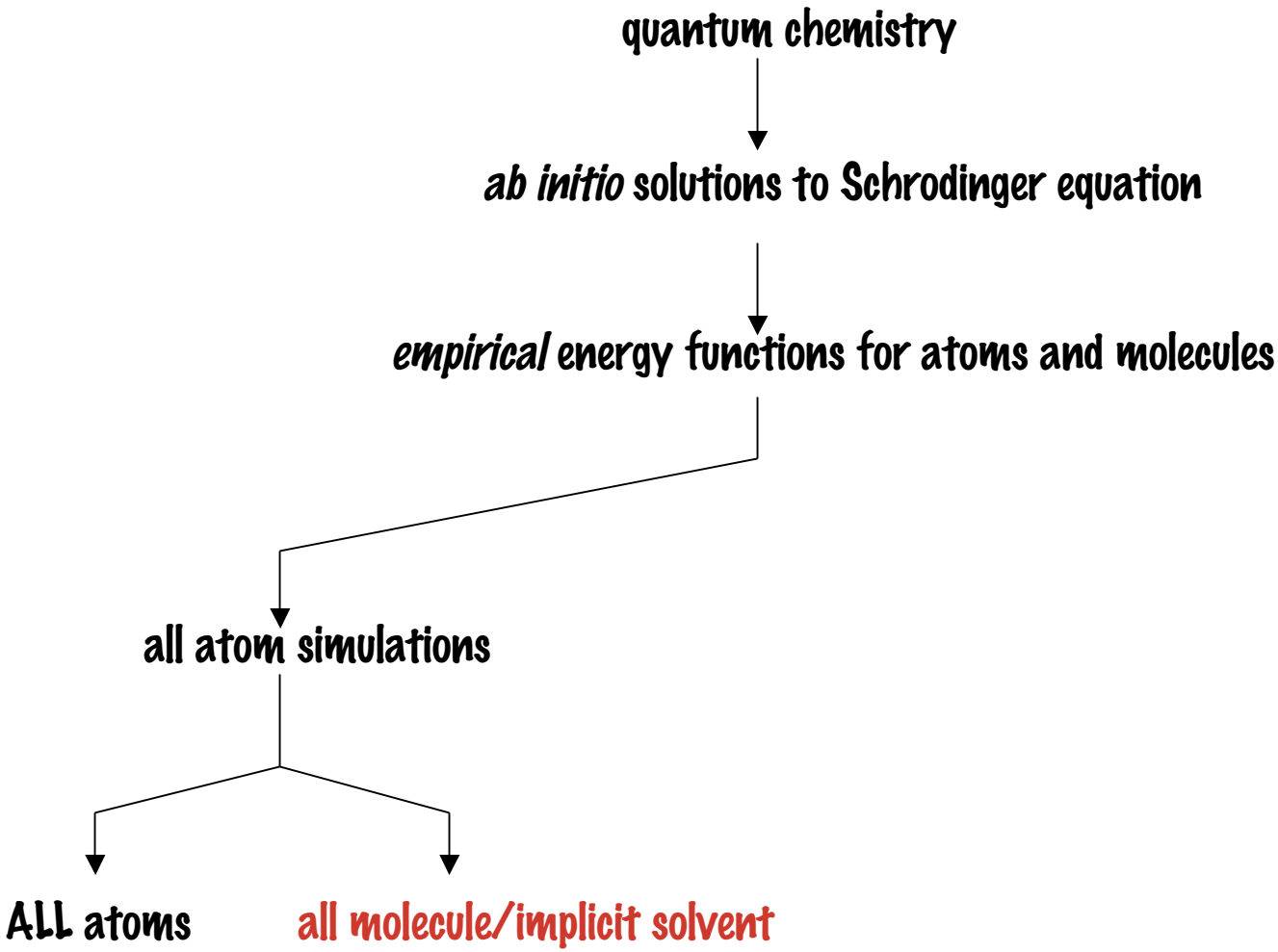
Molecular simulations: Algorithmic and mathematical aspects
Paris 1-3 December 2004

# Levels of representation of peptide: All atom models




* Scheraga, Karplus, Levitt, Kollman and others

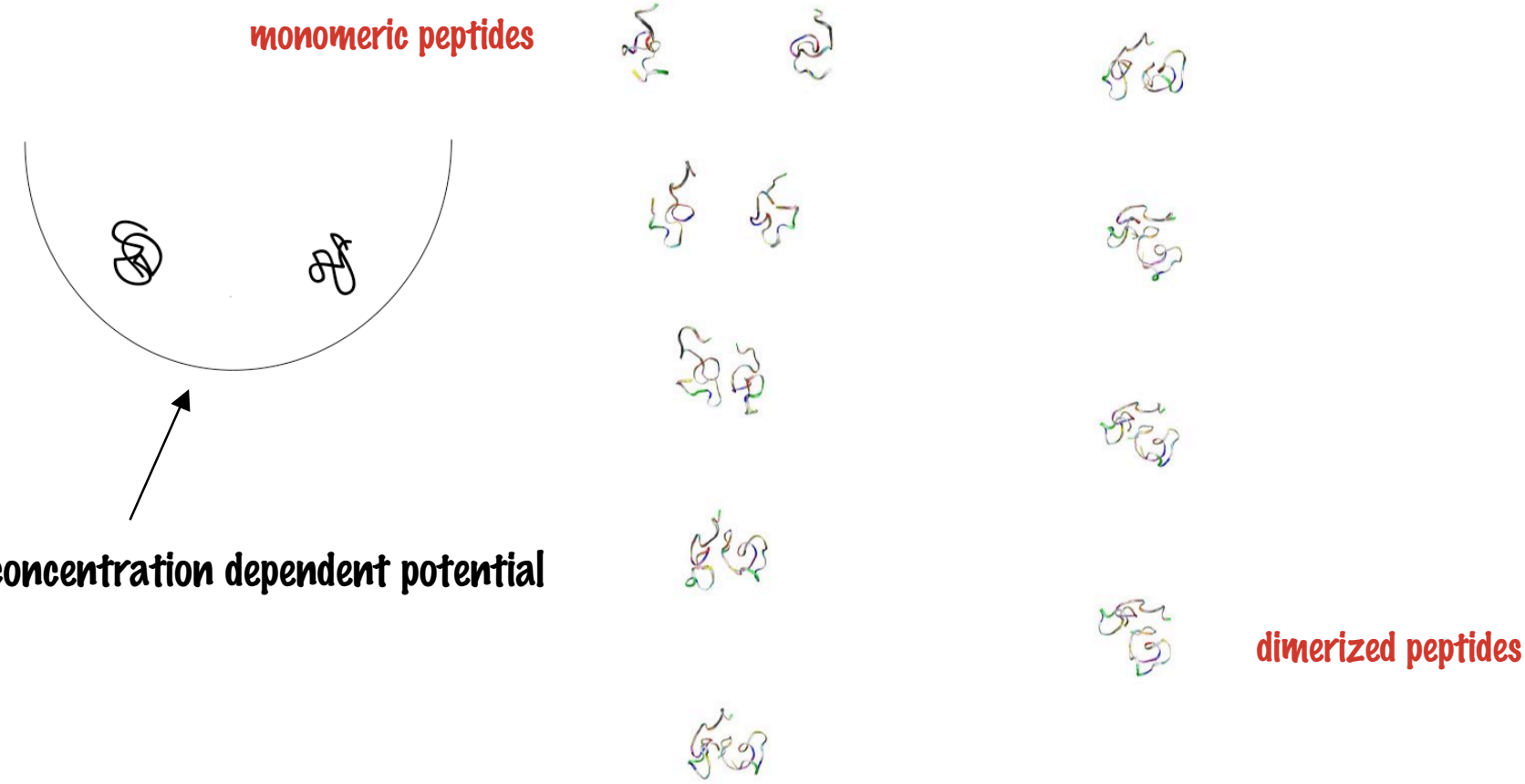
# Levels of representation of peptide: All atom models




* Scheraga, Karplus, Levitt, Kollman and others

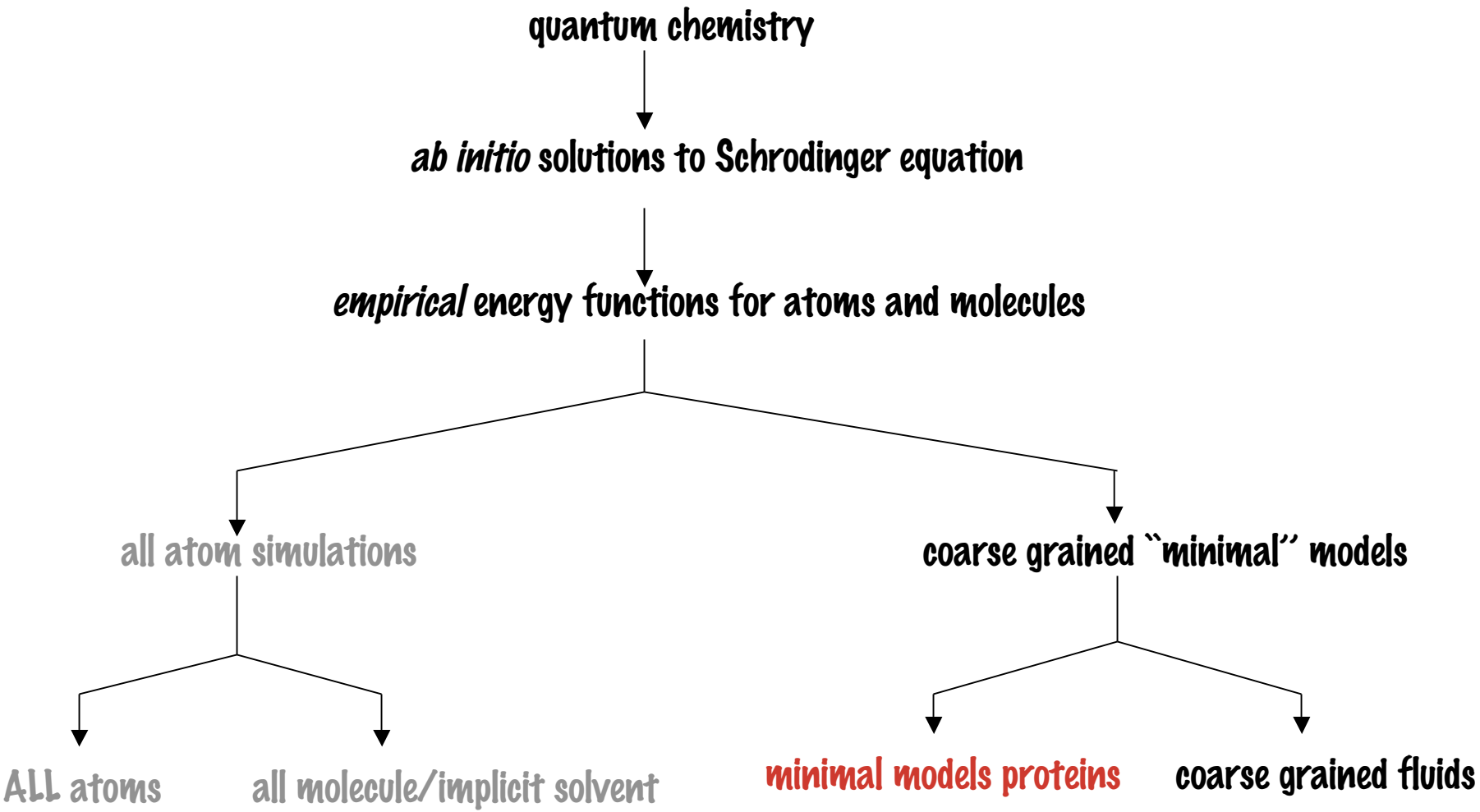
# Simulating dimerization with a concentration dependent potential

- Dimerization simulated using all atoms protein models and IMPLICIT solvation models




- Monomeric collapsed coils dimerize in simulations of 6 x $10^{-4}$ M peptide solution.
- Resulting structures *dominated by ion pairing*, charge-charge contacts.
- Simulations of monomeric peptide alone lead to similar *structural distortions* and *too low* $R_g$

# Levels of representation of peptide: Minimal models




* Scheraga, Levitt and Warshel, Dill, Jernigan, Wolynes and Luthey-Schulten, and others

## Why is there a need for "coarse-grained" potential functions?

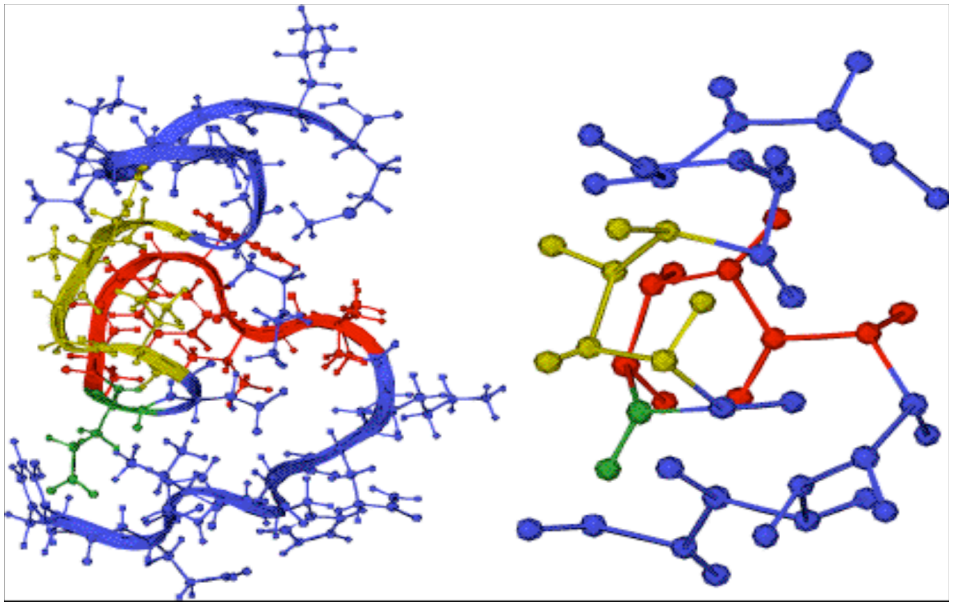
- Rapid molecular structure determination is one of the major goals of *Proteomics*. It is currently impractical to obtain very-high resolution structures on a genome wide scale.

- For *large scale* and/or *long time* molecular simulations, using all-atom fields is inefficient

- *Coarse-grained, low resolution* potentials of interaction may provide a *practical solution* for treating the *complex* problems associated with *protein folding* and *protein functionality* (e.g. recognition).

# Coarse-grained models for protein folding and aggregation

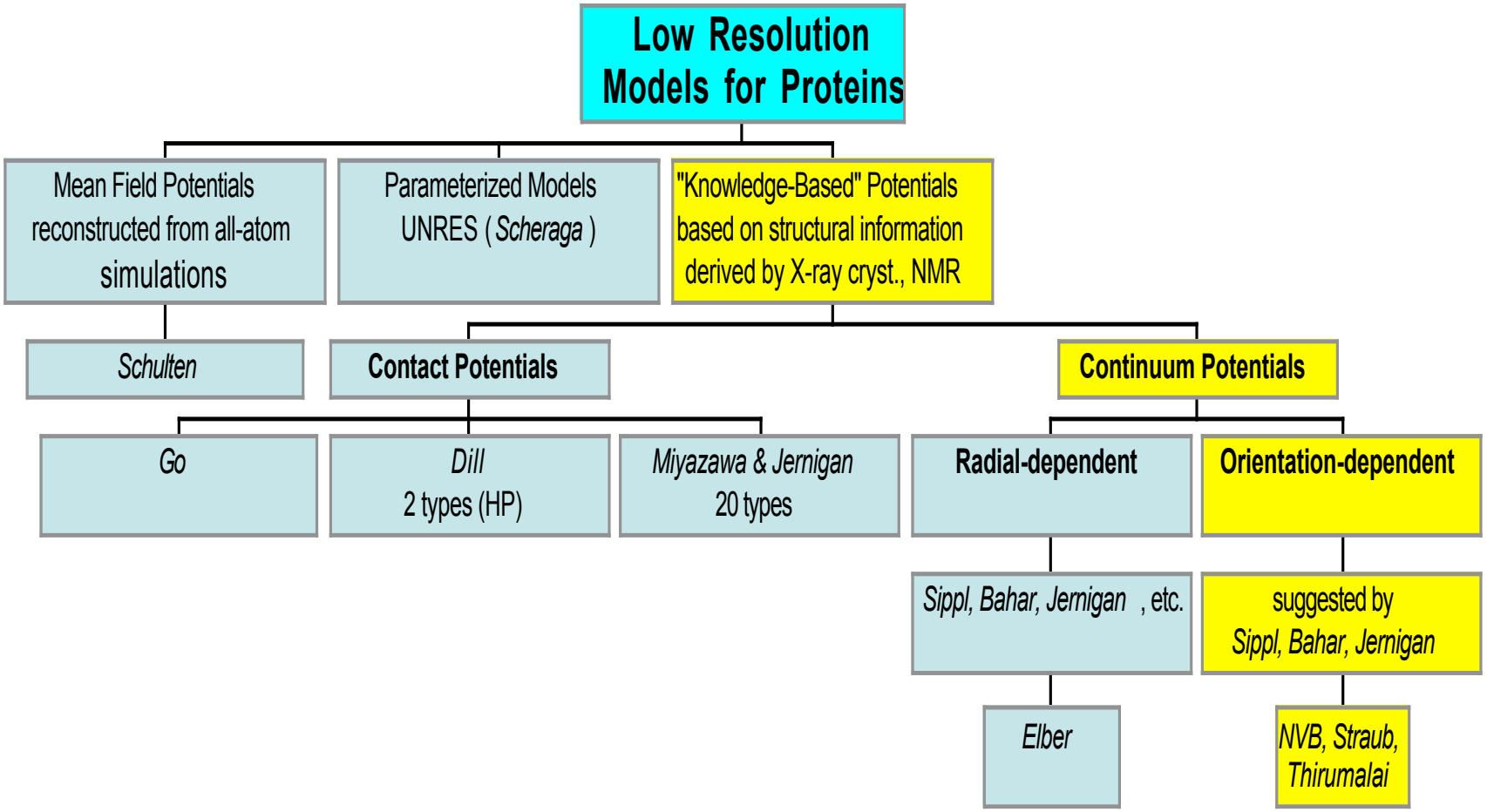
- Reduction from all atoms model to minimal model



•Comparison of energetics of all atom models with statistical models*

•Employ simplified models OFF lattice to explore aggregation thermodynamics#

# A brief history of statistical coarse-grained potentials




- Challenge to build coarse-grained potentials that recognize structures of low free energy.
- Coarse-grained potentials allow for greatly enhanced sampling needed in aggregation studies.

# The "Boltzmann device" connects distributions with potentials

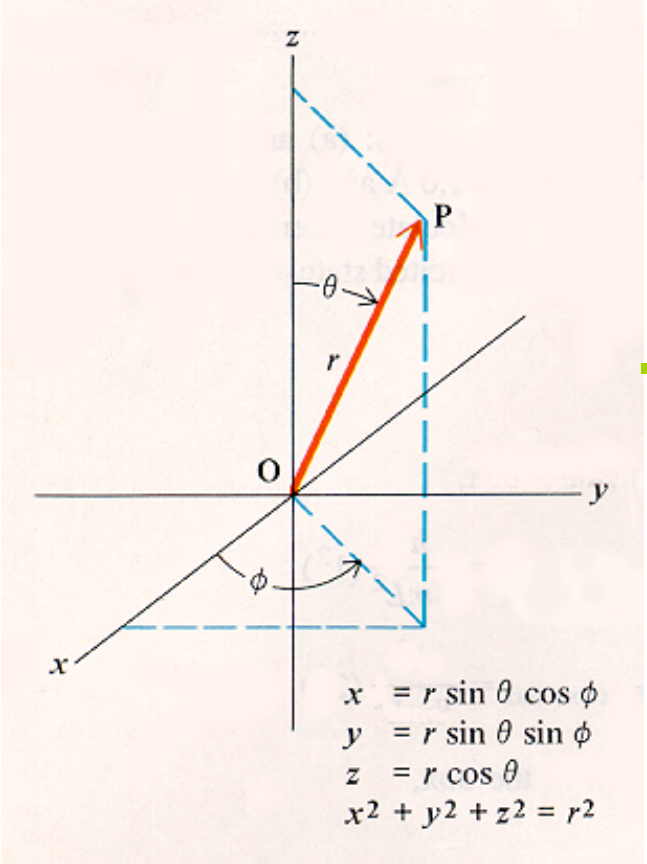
- Sippl (1990) introduced explicit distance dependence in the database-derived potentials of mean force using the Boltzmann formula.
- Basic assumption: known experimentally derived structures from protein databases correspond to *classical equilibrium states*.

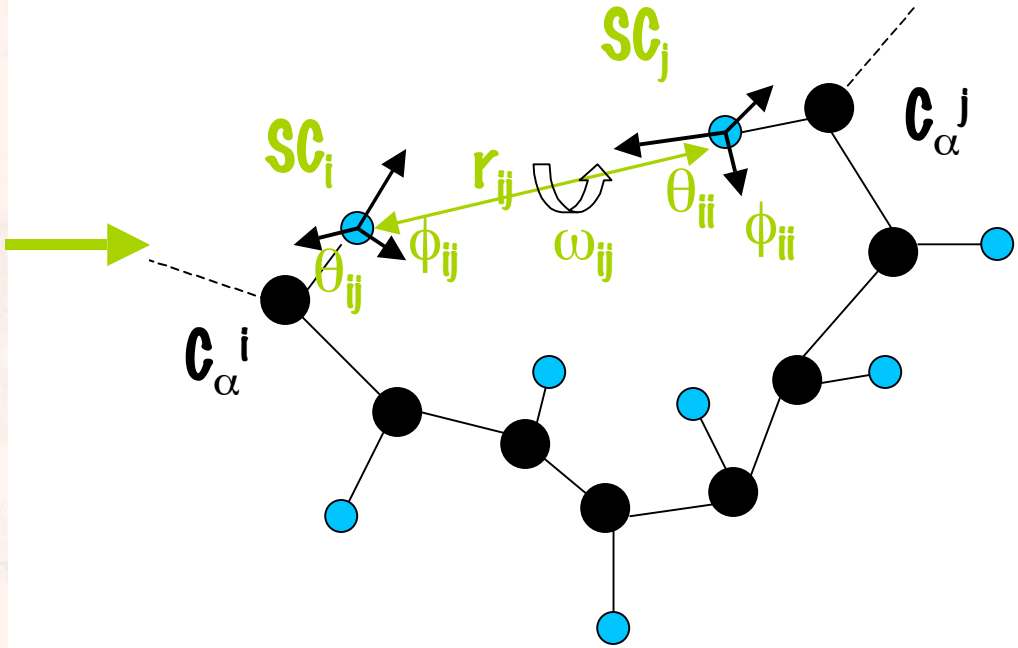
$$U^{ij}(r_{ij}) \propto -kT \ln\left(\frac{P^{ij}(r_{ij})}{P_{ref}(r_{ij})}\right)$$

- The choice of a suitable reference state is very important and it is often the main difference between various potential types

* Miyazawa-Jernigan, Sippl, Levitt

# Reduced Model for Proteins with Local Reference Frames (LRFs)




- More appropriate than Cartesian coordinates for systems with spherical symmetry

# The "Boltzmann device" connects distributions with potentials

- Extension of statistical potentials to include ORIENTATIONAL as well as RADIAL dependence is now possible, due to larger database of protein structures

$$P^{ij}(r_{ij},\phi_{ij},\theta_{ij}) \propto \exp\left(-\frac{U^{ij}_{DO}(r_{ij},\phi_{ij},\theta_{ij})}{kT}\right)$$

- The choice of a suitable reference state is very important and it is often the main difference between various potential types

$$U^{ij}_{DO}(r_{ij},\phi_{ij},\theta_{ij}) = -kT\ln\left(\frac{P^{ij}(r_{ij},\phi_{ij},\theta_{ij})}{P^{ij}_{ref}(r_{ij},\phi_{ij},\theta_{ij})}\right)$$

- Note that $U_{ij} \neq U_{ji}$

* Bahar and Jernigan, Buchete, Straub and Thirumalai

# Make assumption of decomposable probability distributions

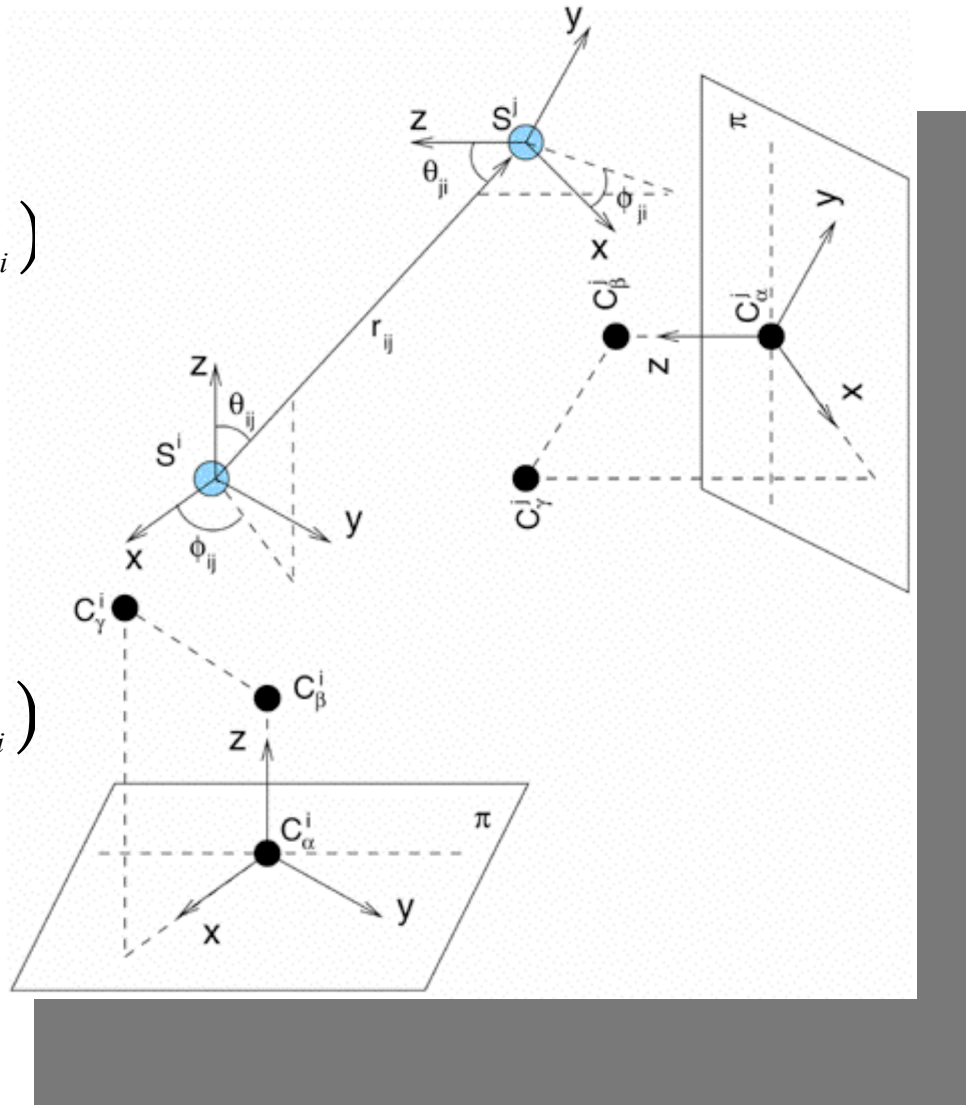
- Quantity of statistics in database necessitates assumption of separable potentials

$$P_{total}^{ij}\left(r_{ij},\varphi_{ij},\theta_{ij},\varphi_{ji},\theta_{ji}\right)=$$
$$P_{total}^{ij}\left(r_{ij},\varphi_{ij},\theta_{ij}\right)\times P_{total}^{ji}\left(r_{ij},\varphi_{ji},\theta_{ji}\right)$$

approximate distribution decomposition

$$U_{DO}^{ij}\left(r_{ij},\varphi_{ij},\theta_{ij},\varphi_{ji},\theta_{ji}\right)=$$
$$U_{DO}^{ij}\left(r_{ij},\varphi_{ij},\theta_{ij}\right)+U_{DO}^{ji}\left(r_{ij},\varphi_{ji},\theta_{ji}\right)$$

leads to additive potential approximation



# Statistical potential for residue/residue interactions in proteins

JOURNAL OF CHEMICAL PHYSICS VOLUME 118, NUMBER 16 22 APRIL 2003

## Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures

N.-V. Buchete and J. E. Straub
*Department of Chemistry, Boston University, Boston, Massachusetts 02215*

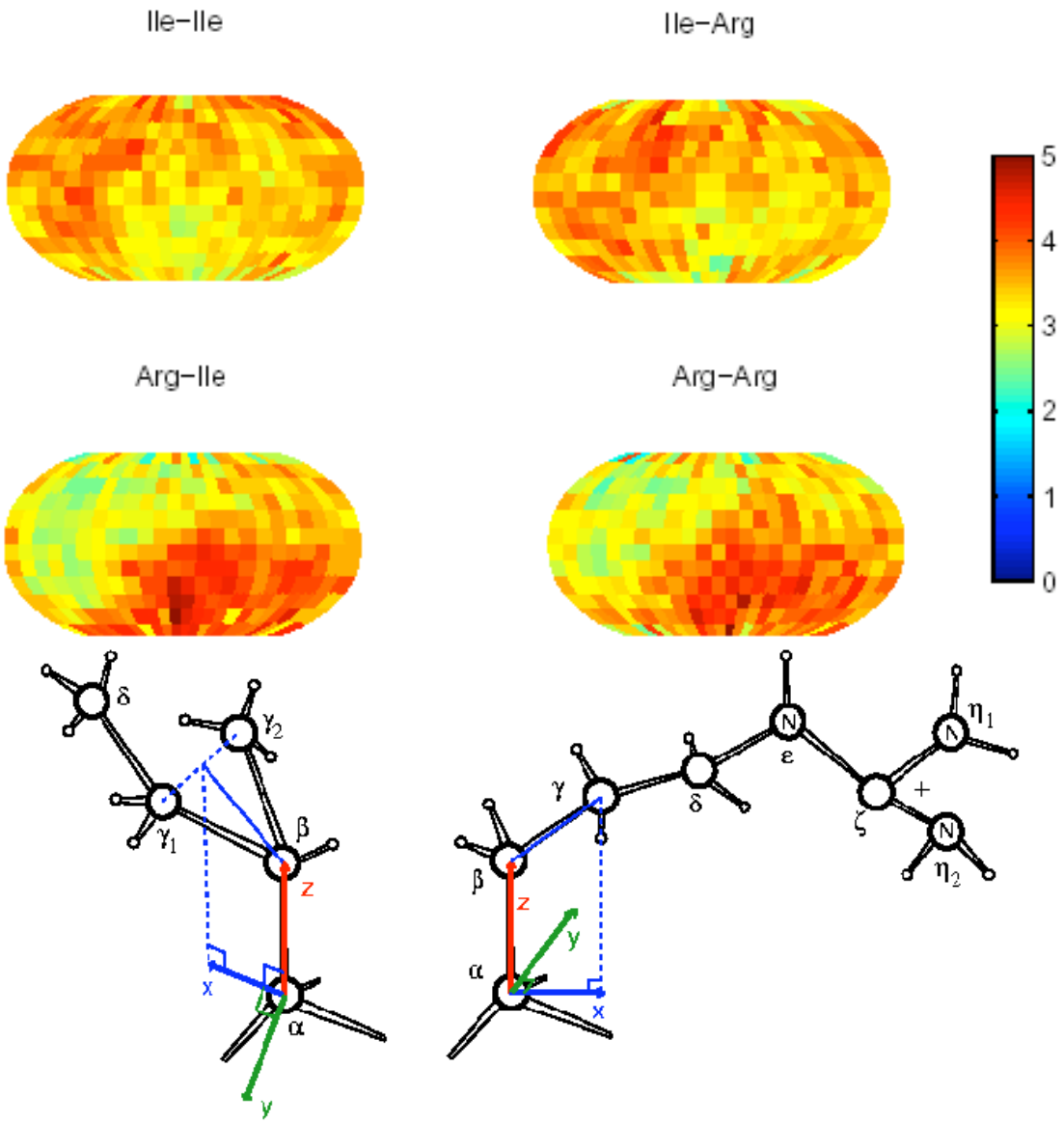
D. Thirumalai
*Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742*

(Received 21 November 2002; accepted 27 January 2003)

We present a new method to extract distance and orientation dependent potentials between amino acid side chains using a database of protein structures and the standard Boltzmann device. The importance of orientation dependent interactions is first established by computing orientational order parameters for proteins with $\alpha$-helical and $\beta$-sheet architecture. Extraction of the anisotropic interactions requires defining local reference frames for each amino acid that uniquely determine the coordinates of the neighboring residues. Using the local reference frames and histograms of the radial and angular correlation functions for a standard set of nonhomologue protein structures, we construct the anisotropic pair potentials. The performance of the orientation dependent potentials was studied using a large database of decoy proteins. The results demonstrate that the new distance and orientation dependent residue–residue potentials present a significantly improved ability to recognize native folds from a set of native and decoy protein structures. © *2003 American Institute of Physics*. [DOI: 10.1063/1.1561616]

Buchete, Straub and Thirumalai, JCP 118, 7658 (2003).

# Relative three-dimensional side chain-side chain orientations

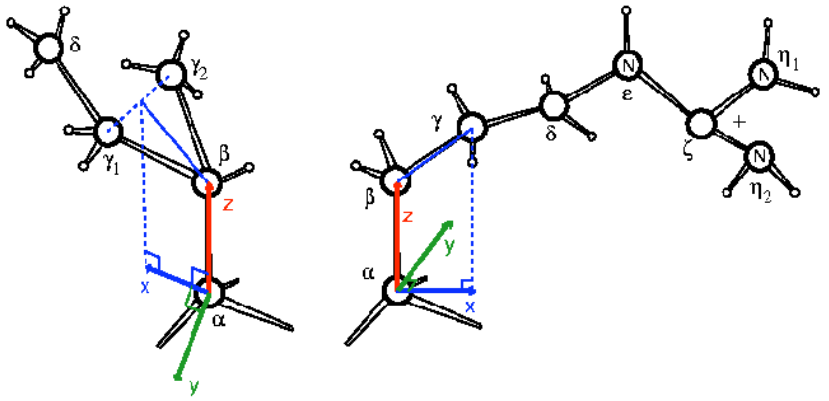


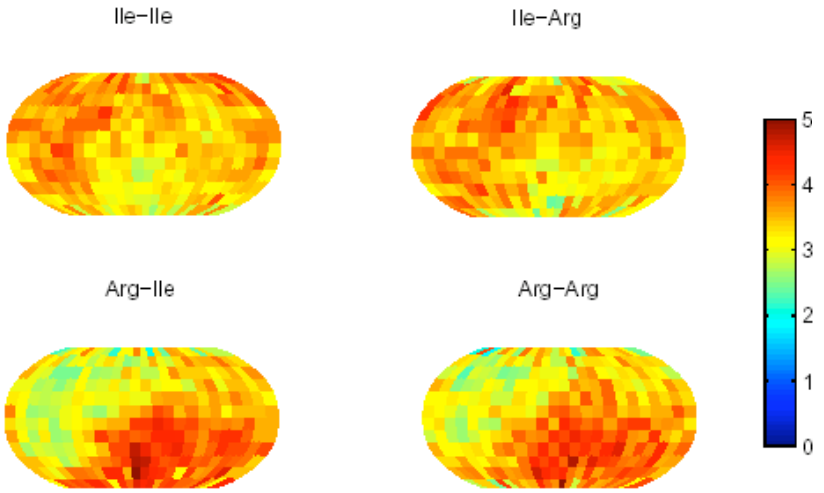

No unique choice! But choice of coordinates is important to overall goodness of potential function

# Interaction potentials derived for 20 amino acids from PDB structures

**Basic assumption: set of *experimentally derived or theoretically generated* structures represent equilibrium distribution of states - *quasichemical assumption.***

$$P^{ij}(r_{ij},\phi_{ij},\theta_{ij}) \propto \exp\left(-\frac{U^{ij}_{DO}(r_{ij},\phi_{ij},\theta_{ij})}{kT}\right)$$



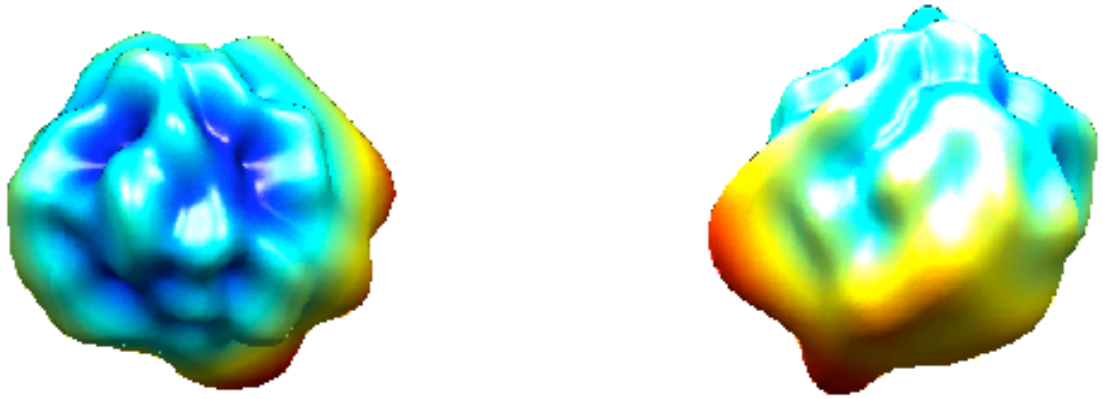




$$U^{ij}_{DO}(r_{ij},\phi_{ij},\theta_{ij}) = -kT\ln\left(\frac{P^{ij}(r_{ij},\phi_{ij},\theta_{ij})}{P^{ij}_{ref}(r_{ij},\phi_{ij},\theta_{ij})}\right)$$

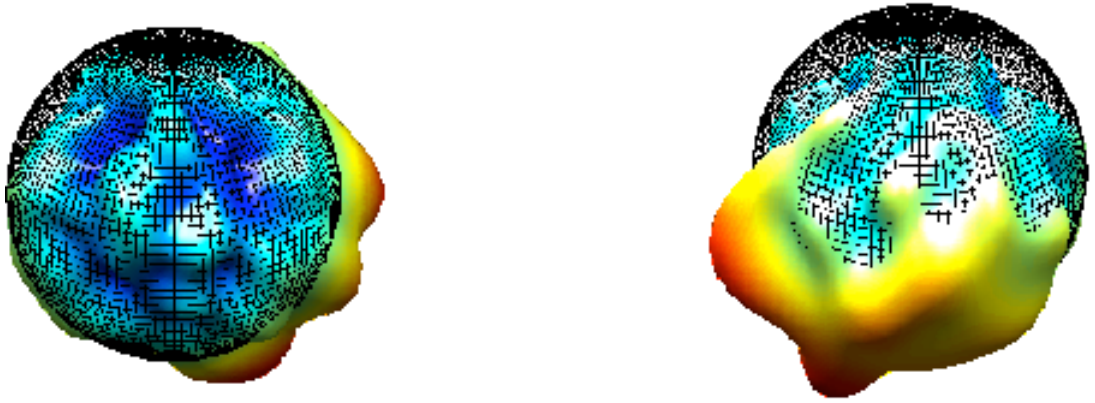
- Miyazawa-Jernigan, Sippl, Levitt, Thirumalai
- Buchete, Straub and Thirumalai, J. Chem. Phys. 118, 7658-7671 (2003); Prot. Sci. 13, 862-874 (2004); Polymers 45, 597-608 (2004); Curr. Opin. Struc. Bio. 14, 225-232 (2004).

# What do our smoothed potentials look like?

- The pattern of spatial anisotropy can be accurately fit with relatively few terms in expansion

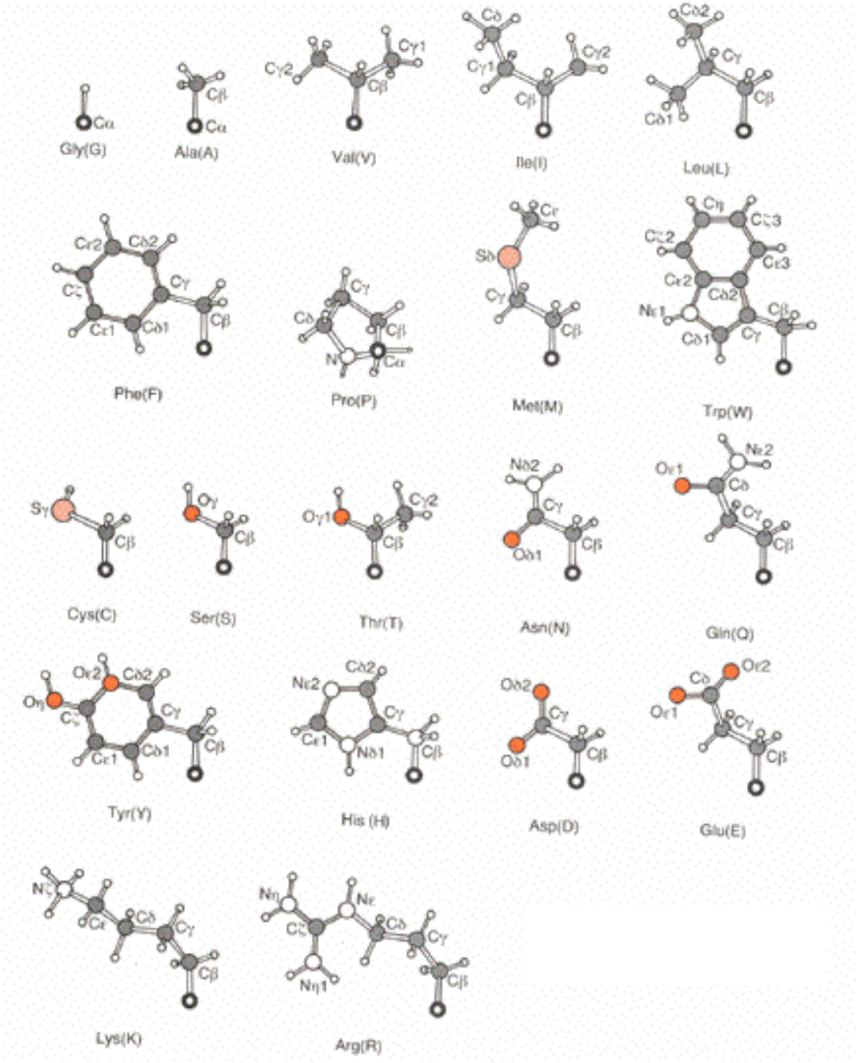


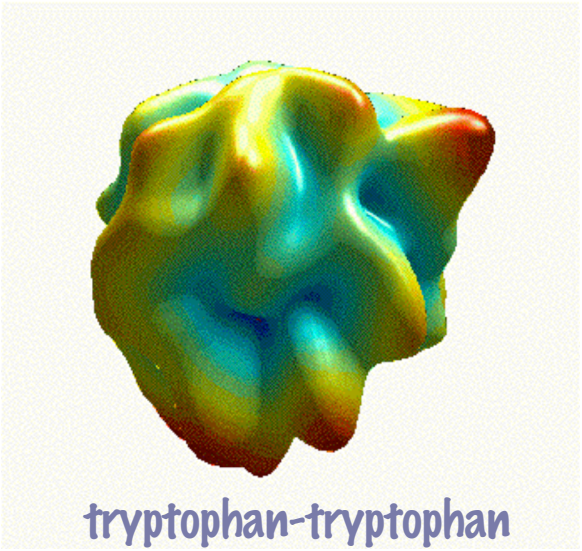
- Substantial spatial aniosotropy essential in capturing excluded volume and polarity for packing



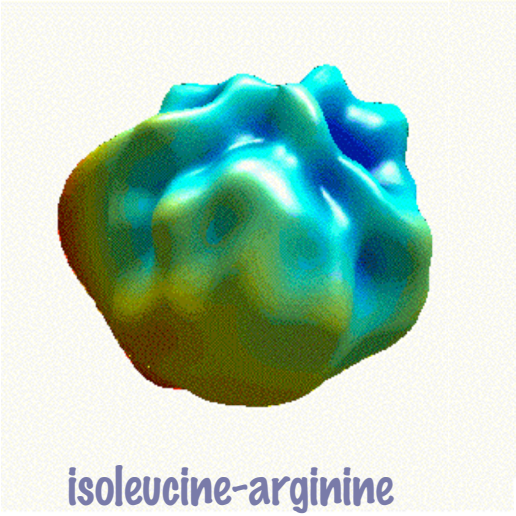
Statistical distributions should capture tendency to form "foldon" motifs for side chain packing

# Smoothed Potentials for use in MC or MD simulations

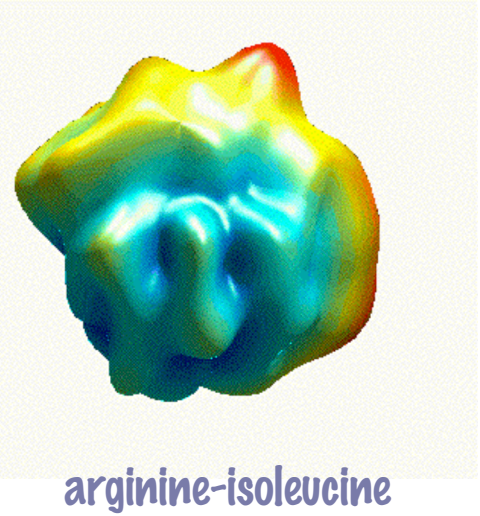


tryptophan-tryptophan



isoleucine-arginine



arginine-isoleucine

Potential appropriate for protein structure *discrimination* given backbone fold topology

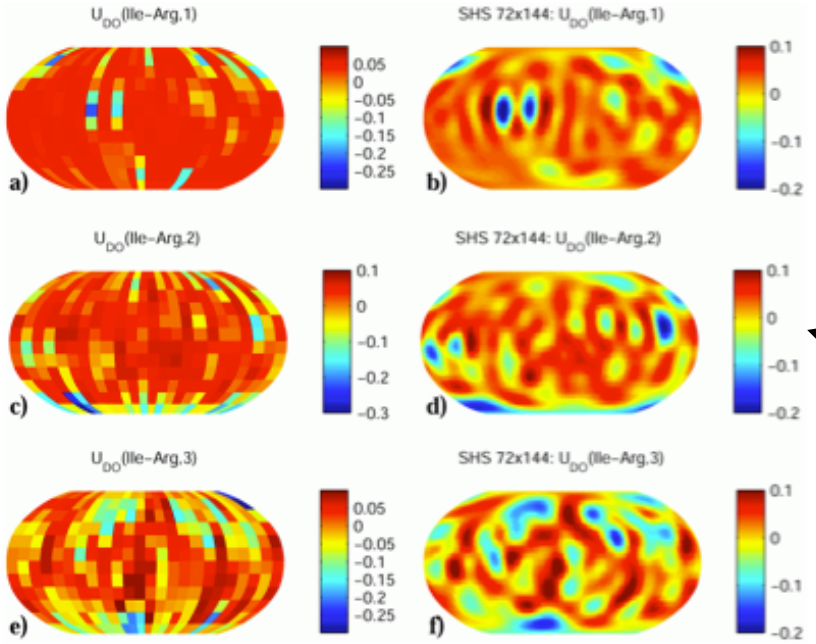
# Spherical Harmonic Synthesis used to fit continuous potentials

The statistical data that is discrete can be fit using "spherical harmonic synthesis" *

$$U(\theta,\phi) = \sum_{n=0}^{N}\sum_{m=0}^{n} P_n^m(\cos\theta)\left(a_{mn}\cos m\varphi + b_{mn}\sin m\varphi\right)$$

$$a_{mn} = \alpha_{mn}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} U(\theta,\varphi)P_n^m(\cos\theta)\cos(m\varphi)\cos\theta\, d\theta d\varphi \qquad b_{mn} = \alpha_{mn}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} U(\theta,\varphi)P_n^m(\cos\theta)\sin(m\varphi)\cos\theta\, d\theta d\varphi$$

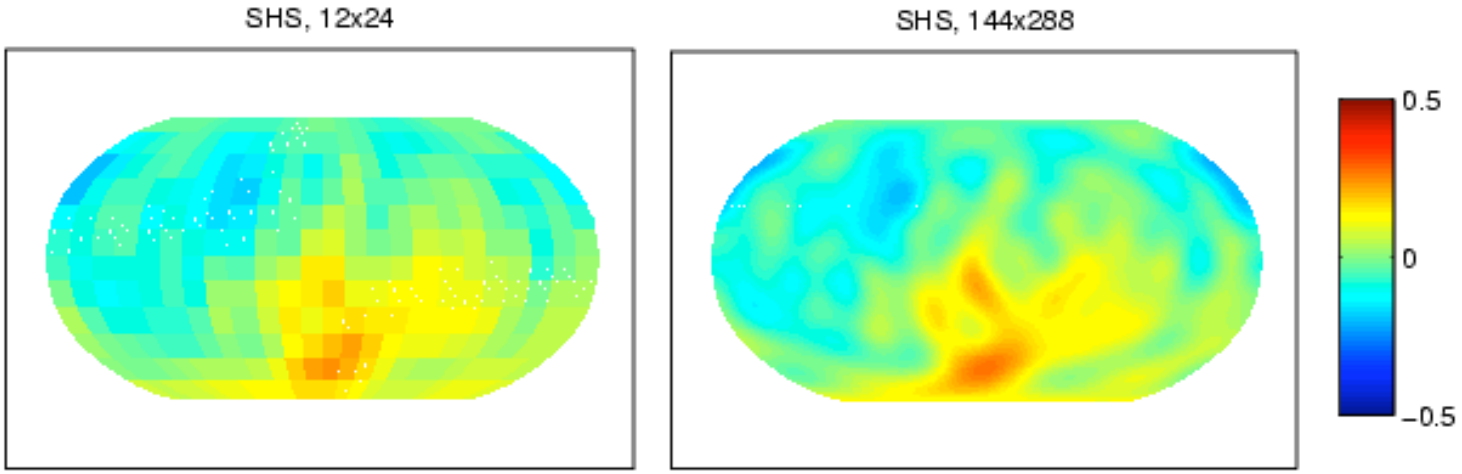
discrete



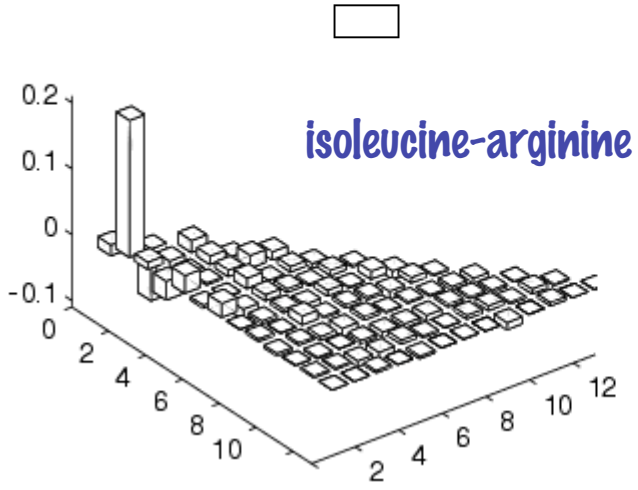

continuous

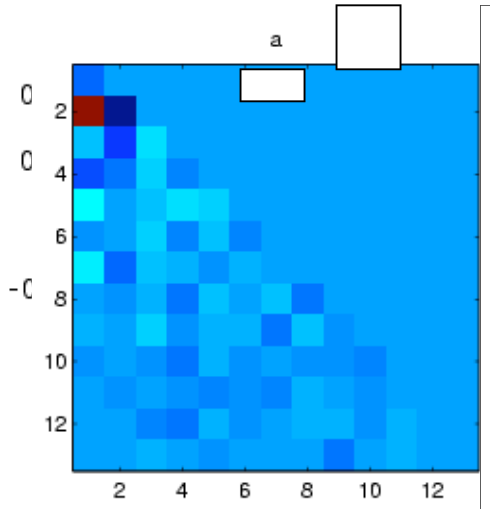
* Adams and Swarztrauber, *Spherepack 3.0*, *MWR*, v.127, p.1872-78 (1999).

# Spherical Harmonic Synthesis provides continuous potentials



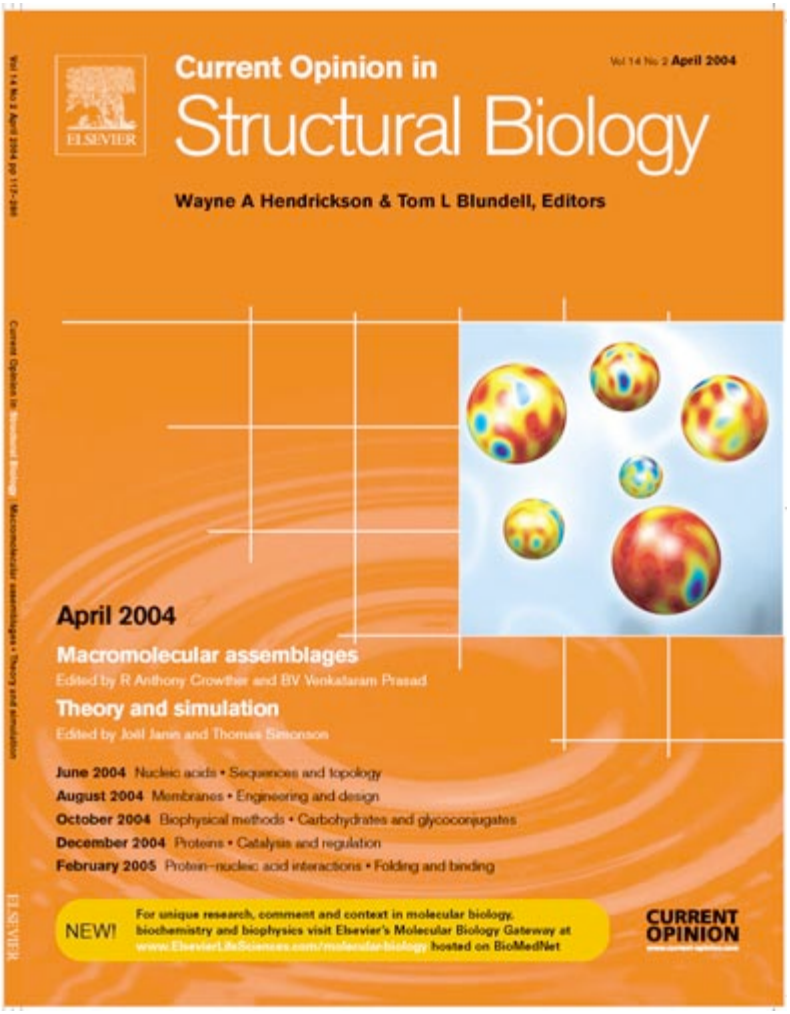

- The distribution of coefficients used in the expansion vary greatly in magnitude




- The pattern of spatial anisotropy can be accurately fit with relatively few terms in expansion

# Different scales ... same (mathematical) nature

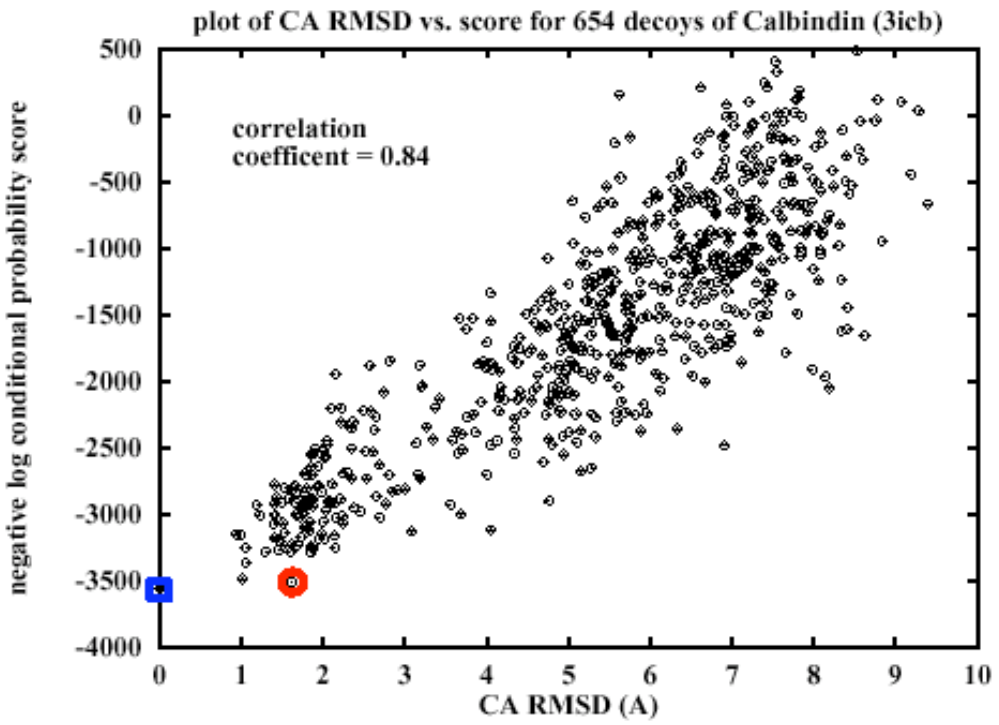



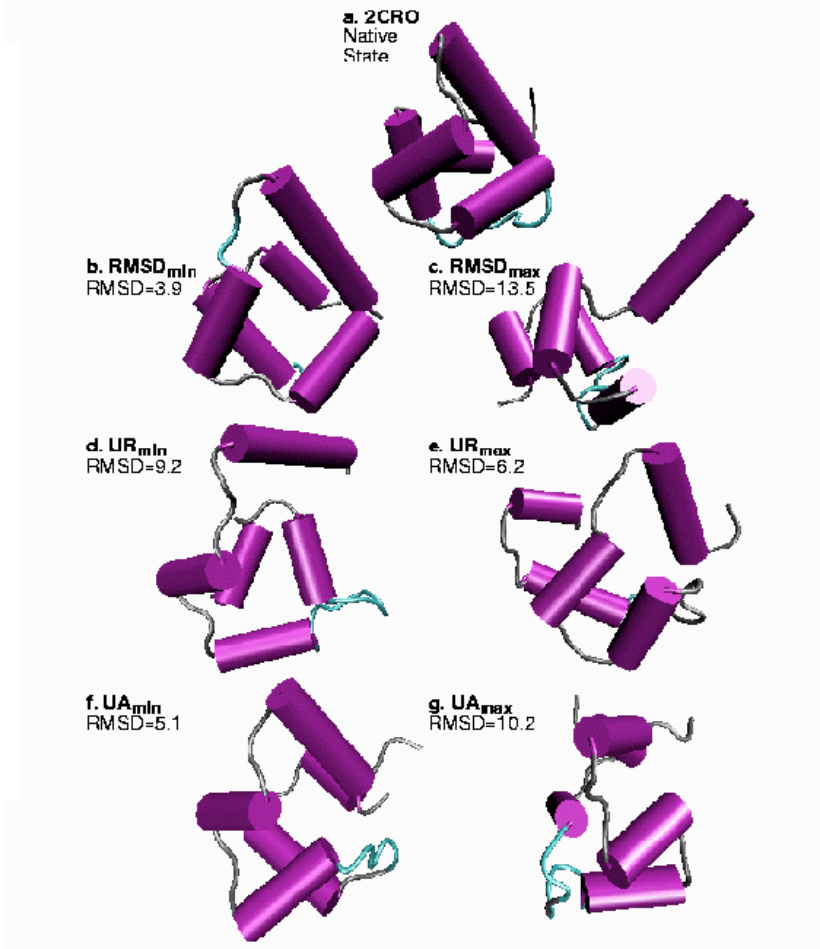

Method used to fit orientation-dependent potentials borrowed from engineering applications.

Buchete, Straub and Thirumalai, *Curr. Opin. Struct. Biol.* 14, 225 (2004).

# Sample Decoys from "Decoys R Us"

- The potential is tested for goodness in recognizing native state in set of decoys



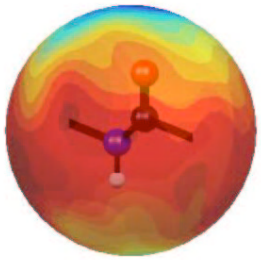



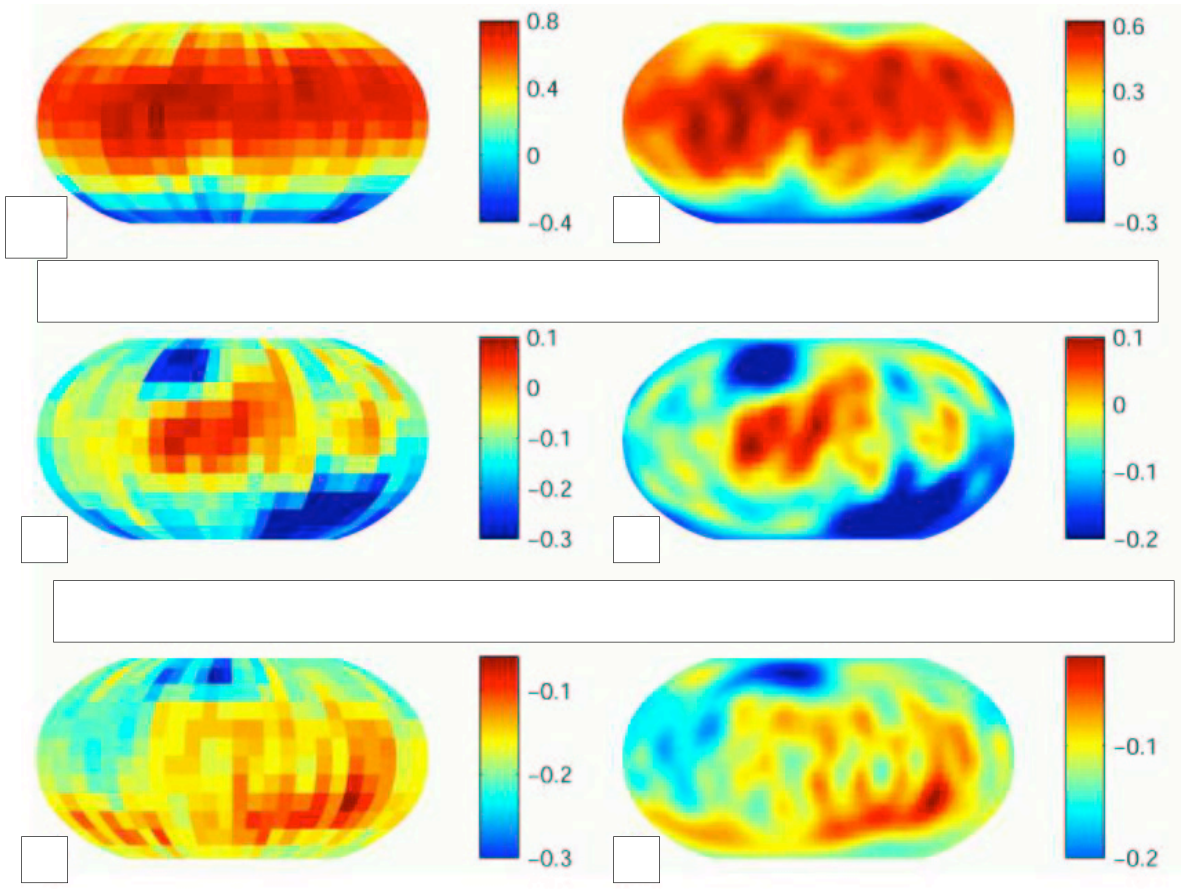

Samudrala R, Levitt M. *Protein Science* (2000), 9:1399-1401 http://dd.stanford.edu

# Statistical potential for backbone-backbone interactions

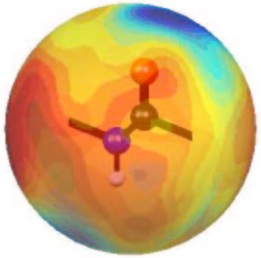
There is a need to add an additional "virtual backbone" site in addition to side chains

short range [2-5.6A]



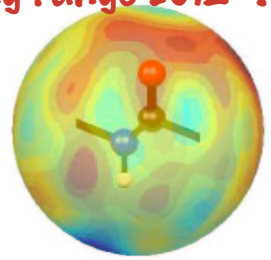



medium range [5.6-9.2A]





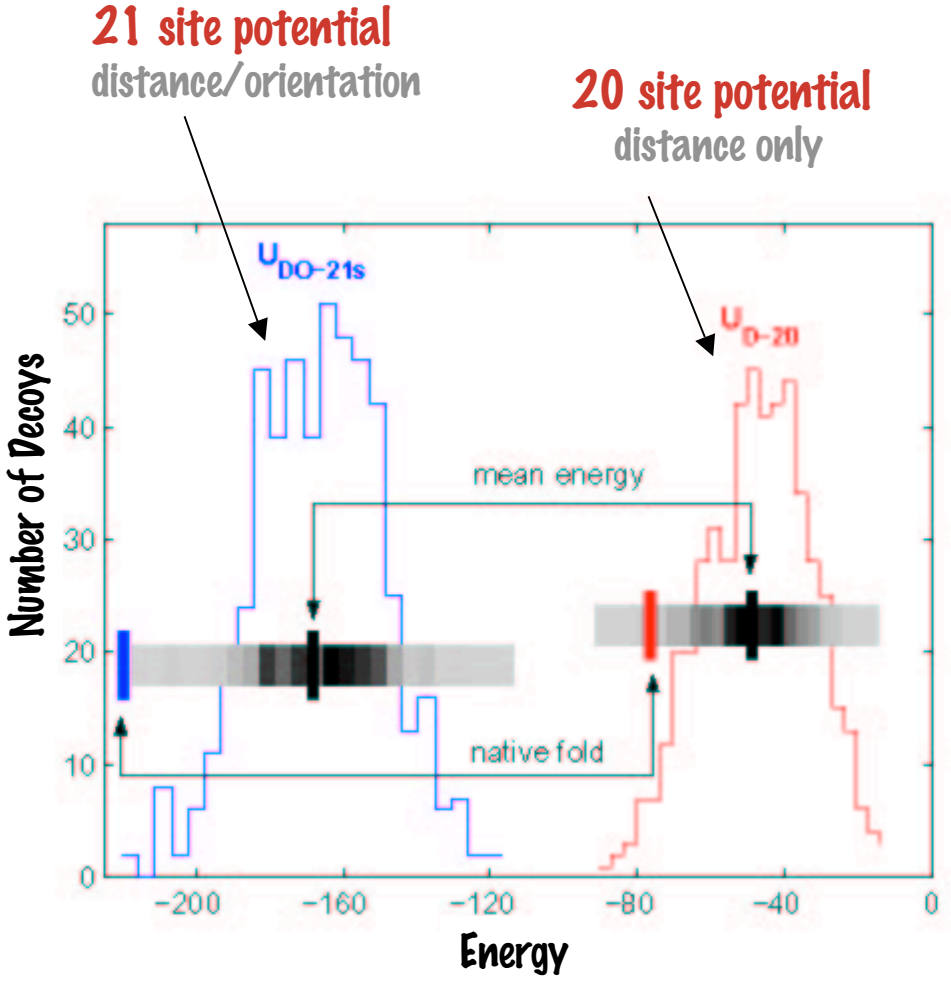

long range [9.2-12.8A]





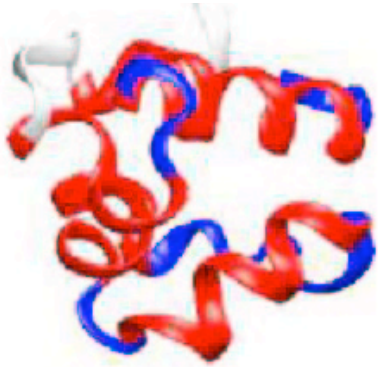

- Buchete, Straub and Thirumalai, Prot. Sci. 13, 862-874 (2004); Polymers 45, 597-608 (2004)

# Great improvement in native state discrimination




native fold

distant decoy



lowest energy
21 site potential

lowest energy
20 site potential

- Buchete, Straub and Thirumalai, J. Chem. Phys. 118, 7658-7671 (2003); Prot. Sci. 13, 862-874 (2004); Polymers 45, 597-608 (2004); Curr. Opin. Struc. Bio. 14, 225-232 (2004).

# Taking the potential off "lattice" for simulations of folding and aggregation

- By varying number of terms in potential expansion, potential can be smoothly varied

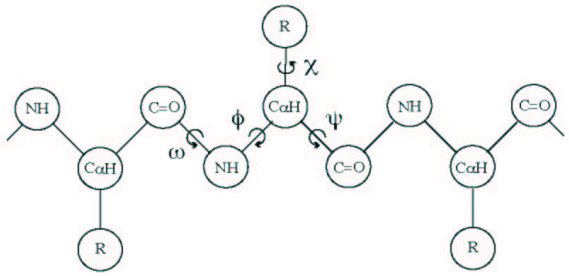
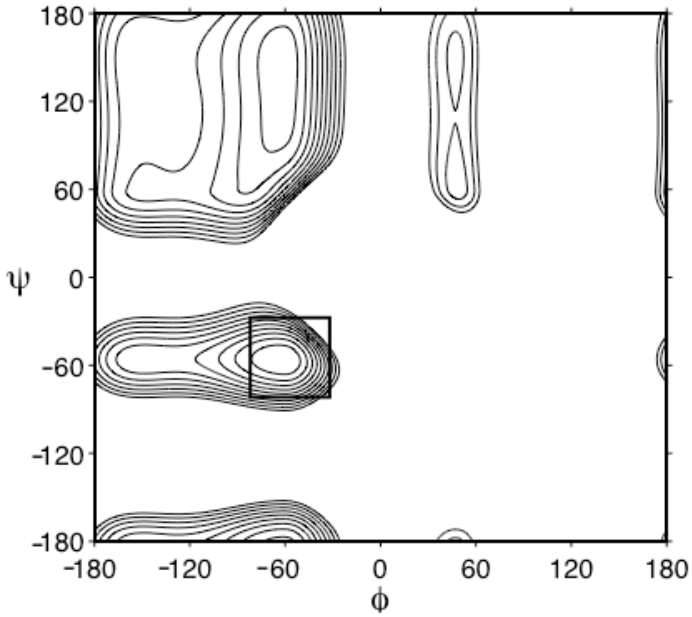


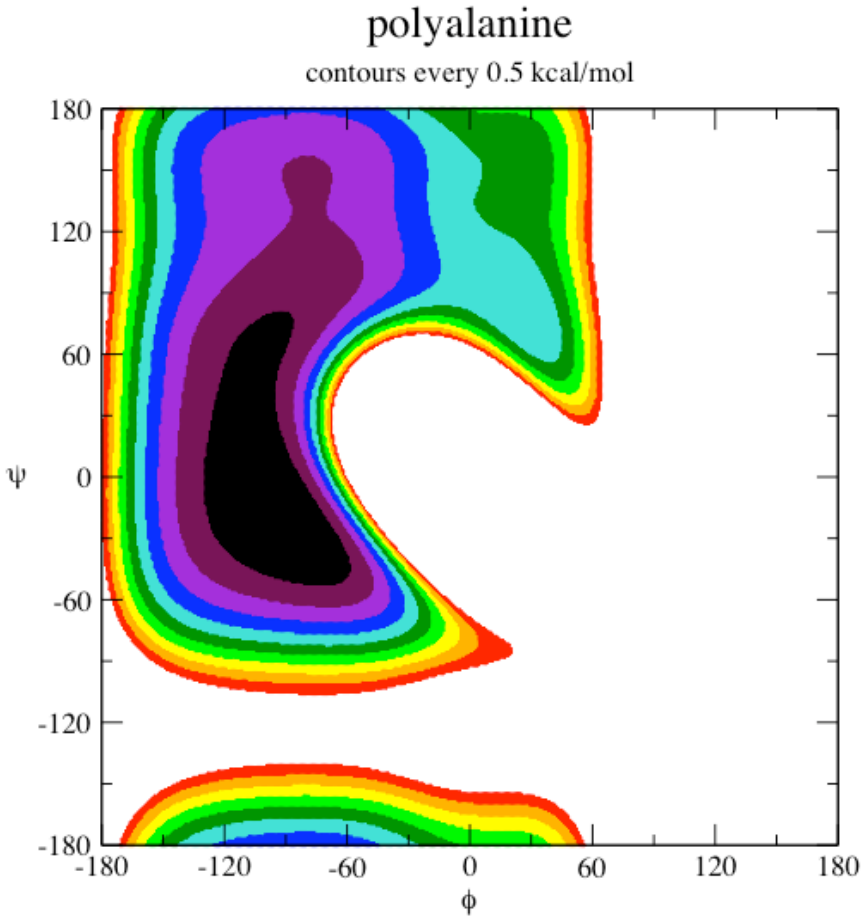

Figure 1:




- How does protein thermodynamics and "dynamics" change as potential is varied?
- Can coarse-graining be carried out directly from a fitting of *effective hamiltonian* to dynamical trajectories?

- Van Giessen and Straub, J. Chem. Phys. (in press, 2004).

# The thermodynamic and kinetic properties depend sensitively on design



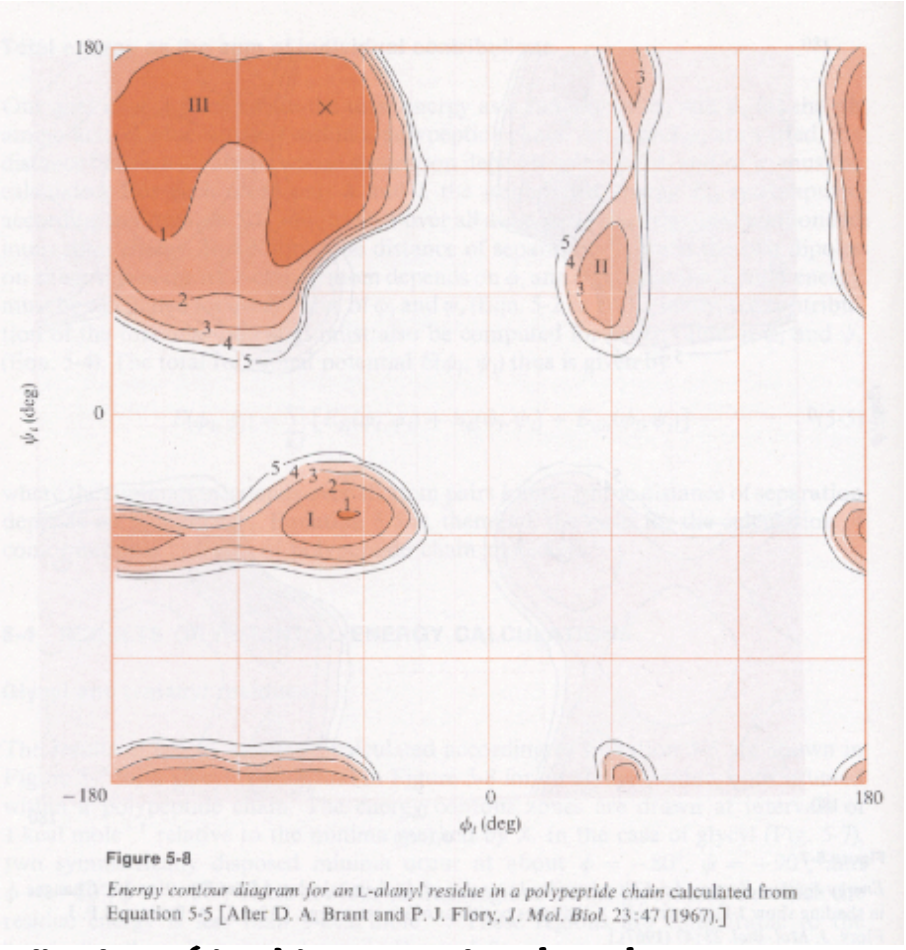



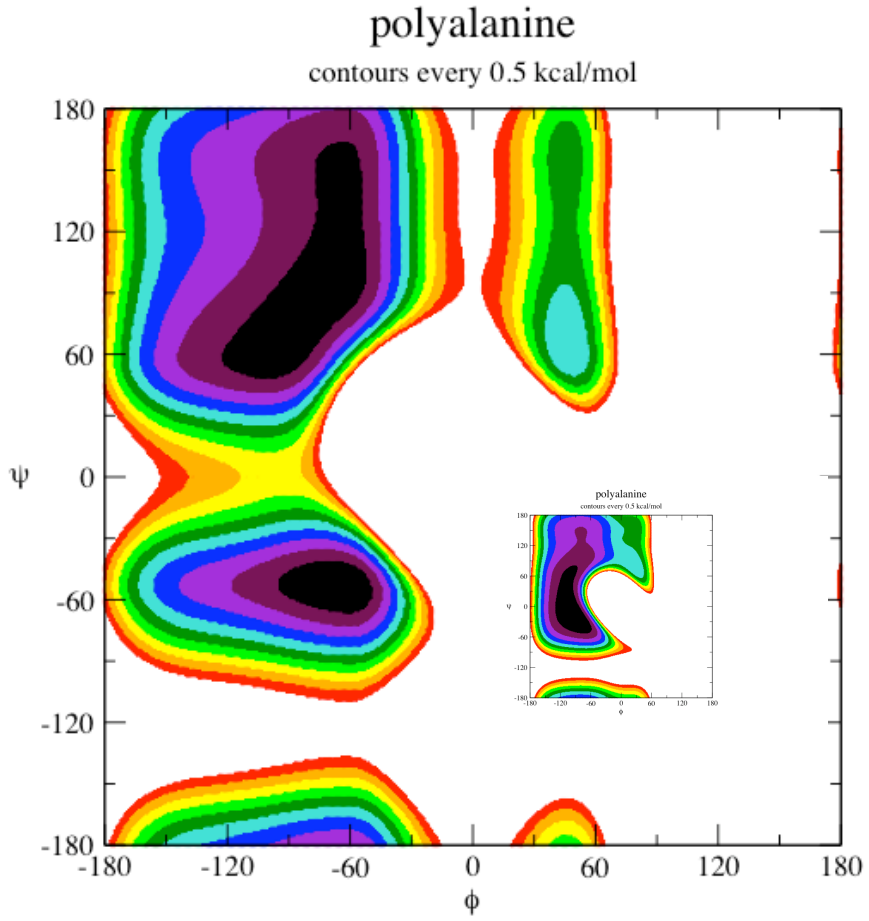

Figure 5-8

*Energy contour diagram for an L-alanyl residue in a polypeptide chain* calculated from Equation 5-5 [After D. A. Brant and P. J. Flory, *J. Mol. Biol.* 23:47 (1967).]

An apparently reasonable estimate of the distribution of backbone torsional angles

S. Takada, Z. Luthey-Schulten, and P.G. Wolynes, *J. Chem. Phys.* **110**, 11616 (1999).

# The thermodynamic and kinetic properties depend sensitively on design



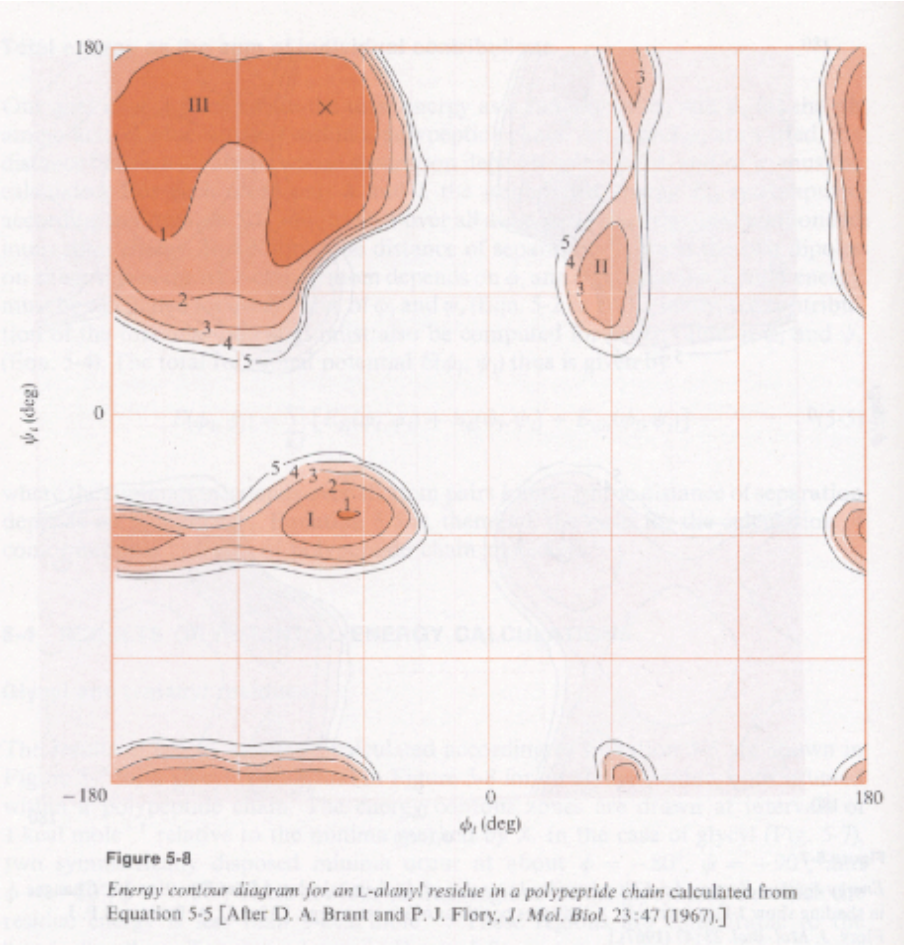




**Figure 5-8**

*Energy contour diagram for an L-alanyl residue in a polypeptide chain* calculated from Equation 5-5 [After D. A. Brant and P. J. Flory, *J. Mol. Biol.* 23:47 (1967).]

Change in backbone potential leads to 200K shift in folding transition and 100x in transition time!

# Systematic variation in coarse-grained model - how is dynamics influenced?

There are three terms in the potential energy

$$V = \lambda V_{SHS} + V_{vdW} + V_{tor}$$

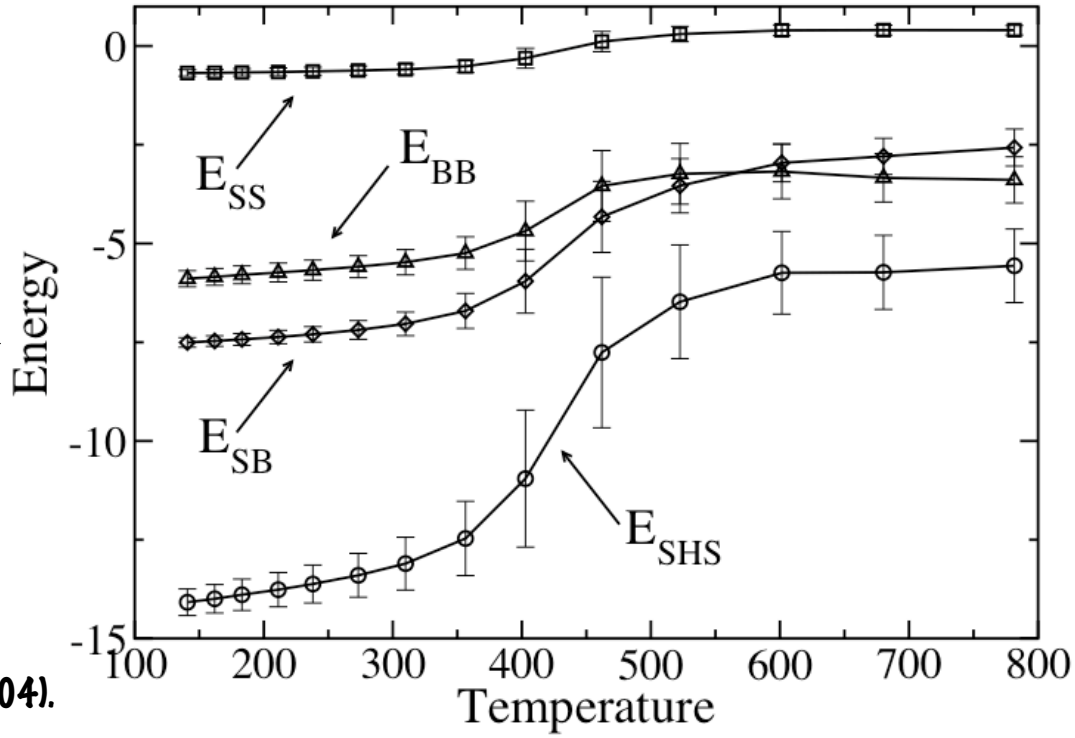
statistics-based interaction energy

van der Waals energy

dihedral angle potential.

How should the relative strengths of the interactions be selected? What is $\lambda$?

relative interactions coil-to-helix in $Ala_{16}$




Van Giessen and Straub, J. Chem. Phys. (in press, 2004).

# Monte Carlo simulation methodology

- The thermodynamic properties are studied using Replica Exchange Monte Carlo

  14 walkers at a range of temperatures from 140 to 780K were used.

- Moves were accepted with the probability

$$P_{\text{accept}} = \min\left(1, \frac{W(\vec{\varphi}' \rightarrow \vec{\varphi})}{W(\vec{\varphi} \rightarrow \vec{\varphi}')} \exp\{-(E' - E)/kT\}\right)$$

where $W(\varphi' \rightarrow \varphi)/W(\varphi \rightarrow \varphi')$ is included to satisfy the demands of detailed balance.

- There are two types of moves in the MC move set

  1) A PIVOT move

  2) A CONCERTED ROTATION-like move [which is faster and computationally less complex than a true concerted-rotation move]

# Monte Carlo simulation methodology (continued)

The PIVOT move consists of rotating the $\varphi$ and $\psi$ angles of one or two residues (that are within 6 residues of each other).

Schematically, the move is

$$\varphi \rightarrow \varphi' = \varphi + d\varphi$$

where $d\varphi$ is drawn from Gaussian distribution with variance of 4 degrees *centered on zero.*

To improve inter-basin crossing, $d\varphi$ is occasionally *centered on 120°.*

# Monte Carlo simulation methodology (continued)

The CONCERTED ROTATION-like move* updates 8 consecutive dihedral angles so as to keep both ends of the polypeptide chain approximately *fixed in space.*

Four residues are chosen (*k*, *k*+1, *k*+2, and *k*+3) as well as three particles in residues *k*+3 (C$\alpha$ and C) and *k*+4 (N) which will remain fixed in space. We define the quantity $\Delta$ so that

$$\Delta^2 = \sum_{I=1}^{3} (\delta \vec{r}_I)^2$$

$r_I$ vectors to each fixed atom

where $\Delta^2 \approx \sum_{i,j,=1}^{\nu} \delta\varphi_i G_{ij} \delta\varphi_j$ and the matrix G has the elements $G_{ij} = \sum_{I=1}^{3} \frac{\partial \vec{r}_I}{\partial \varphi_i} \cdot \frac{\partial \vec{r}_I}{\partial \varphi_j}$

The new angles $\varphi$ are then drawn from the distribution

$$W(\vec{\varphi} \rightarrow \vec{\varphi}') = \frac{1}{\pi^3} (\det \mathbf{A})^{1/2} \exp\left\{ -(\vec{\varphi} - \vec{\varphi}')^T \mathbf{A} \, (\vec{\varphi} - \vec{\varphi}') \right\}$$

where

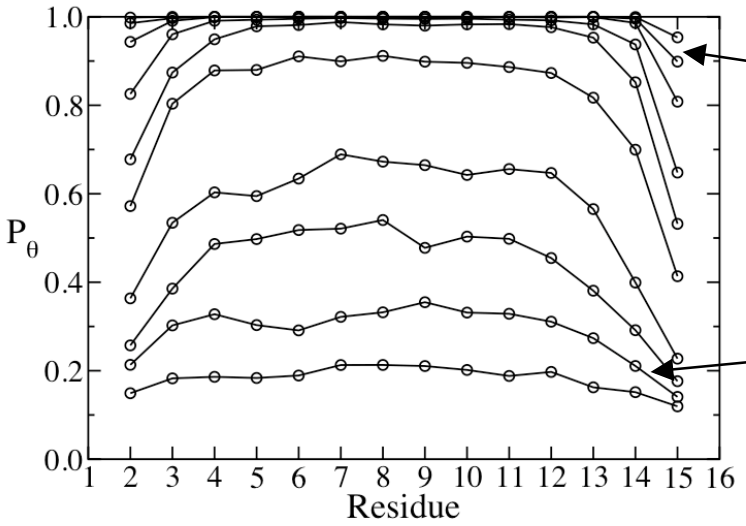
$$\mathbf{A} = \frac{a}{2} (\mathbf{1} + b \, \mathbf{G})$$

Large "*a*" decreases step size; large values of "*b*" gives local moves; *b*=0 gives "random" moves.

* G. Favrin, A. Irbäck, and F. Sjunnesson, *J. Chem. Phys.* **114**, 8154-8158 (2001).

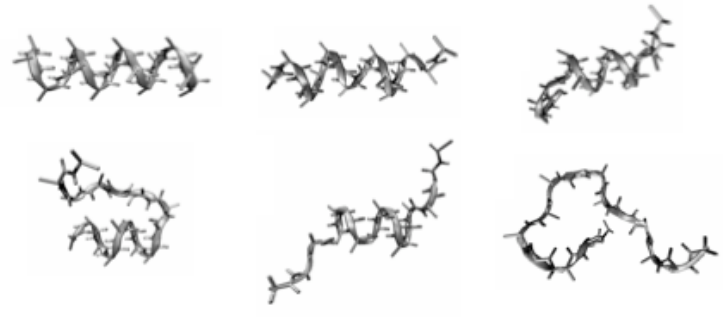
# Coil-to-helix transition in polyalanine peptides

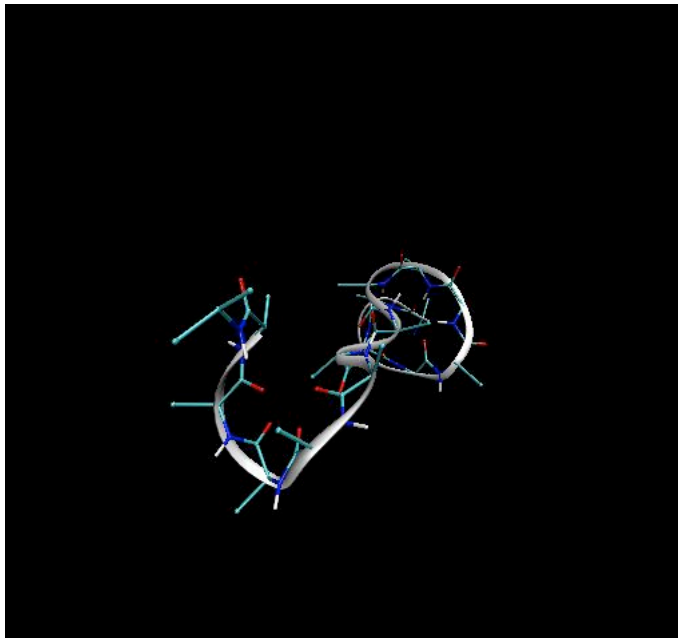
Sampling efficiently performed using replica exchange Monte Carlo method



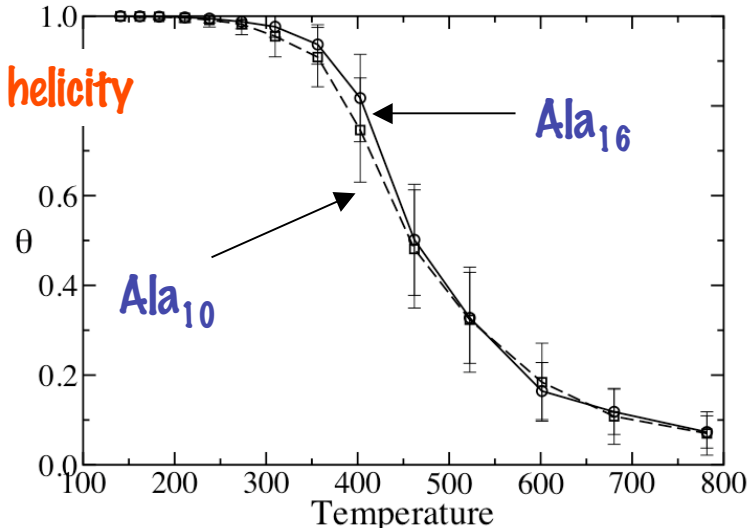

LOW temperature mostly helical

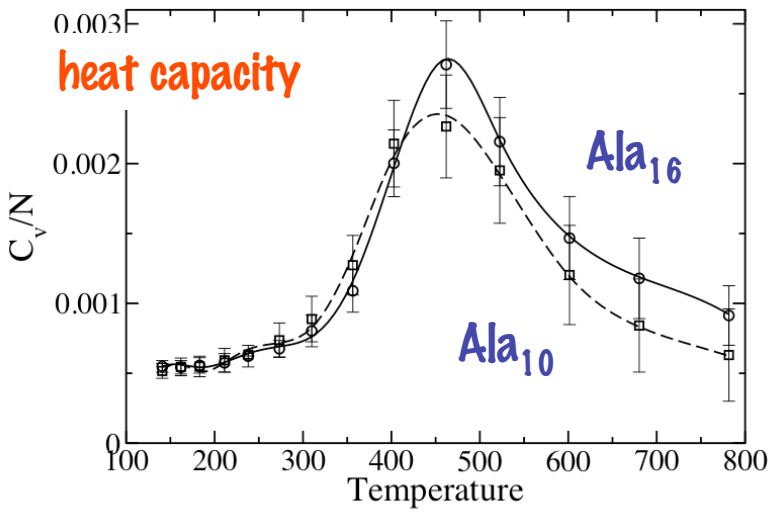
HIGH temperature mostly helical

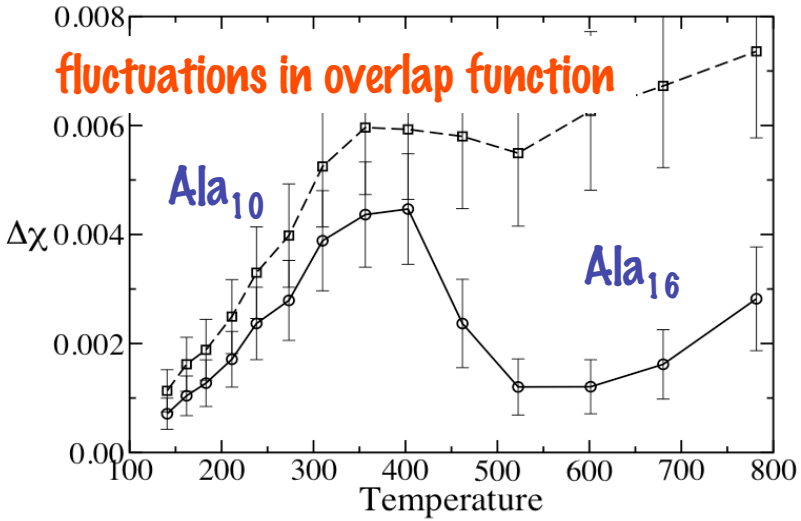


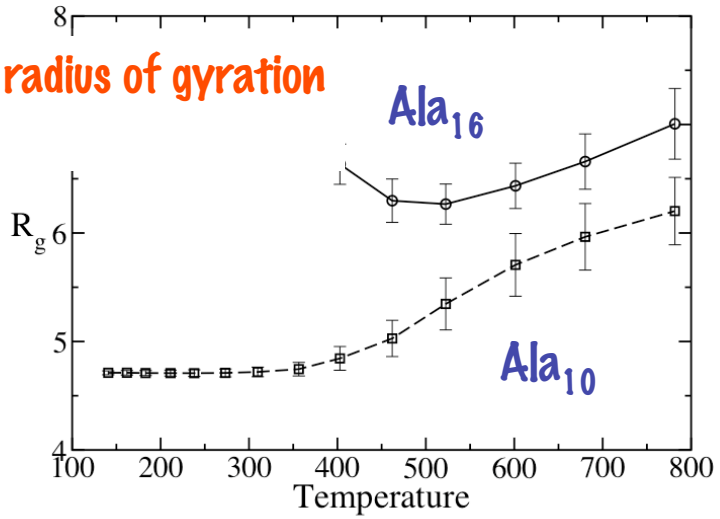


# Coil-to-helix transition in polyalanine peptides



# Coil-to-helix transition in polyalanine peptides

The Zimm-Bragg theory is exactly solvable Ising-like model for the coil-to-helix transition

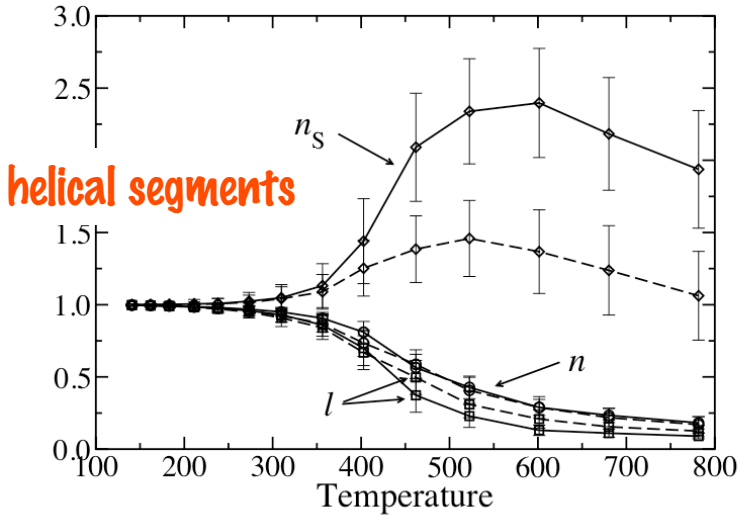
$$\frac{n_H}{N} = \frac{1}{2} - \frac{1-s}{2\sqrt{(1-s)^2 + 4s\sigma}}$$

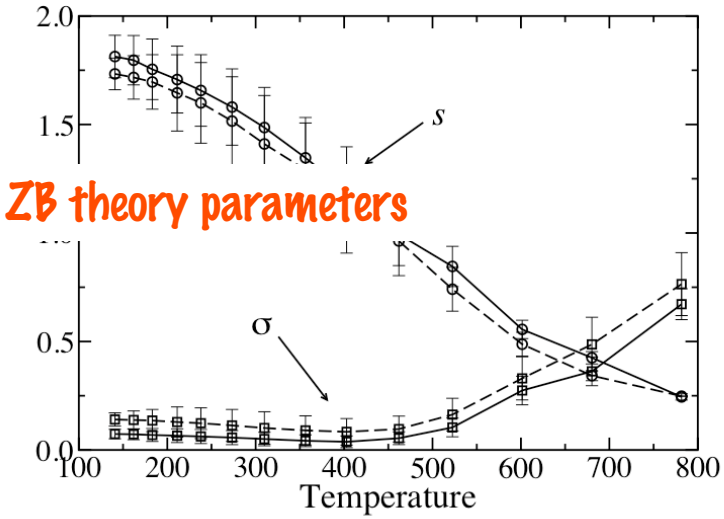
average helix length

$$l = \frac{\langle n_H \rangle}{\langle n_S \rangle} = 1 - \frac{2s}{1-s+\sqrt{(1-s)^2 + 4s\sigma}}$$

average helical segment length

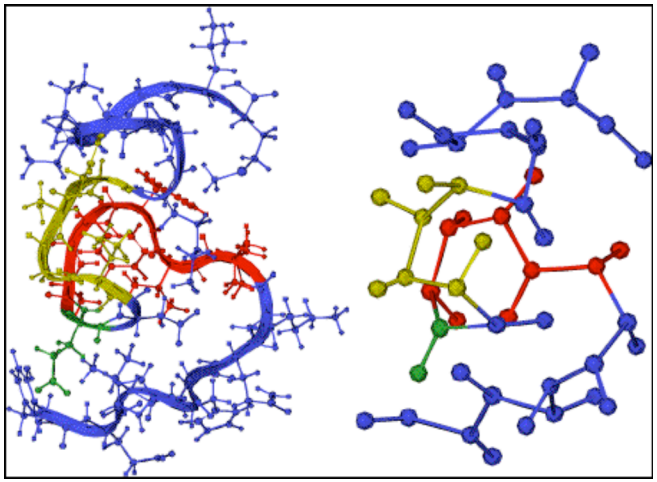
Values of s > 1 indicate strong *propagation* of helix; σ gives tendency for helix *nucleation*

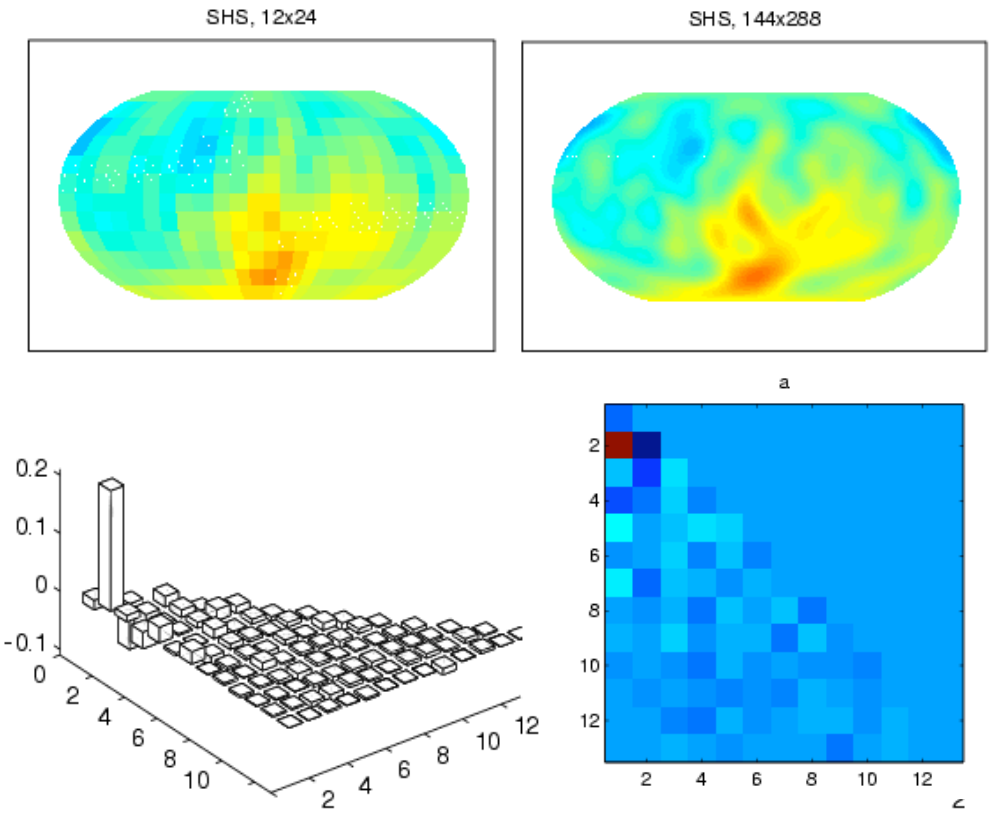


# Systematic variation in coarse-grained model - how is dynamics influenced?

- By varying number of terms in potential expansion, potential can be smoothly varied






- How does protein thermodynamics and "dynamics" change as potential is varied?
- Can coarse-graining be carried out directly from a fitting of *effective hamiltonian* to dynamical trajectories?

- Seminal work of Scheraga, Levitt and Warshel

# Systematic variation in coarse-grained model - how is dynamics influenced?

- By varying number of terms in potential expansion, potential can be *smoothly* varied
- How do protein thermodynamics and "dynamics" change as potential is varied?

$$U(\theta,\phi) = \sum_{n=0}^{N} \sum_{m=0}^{n\ \prime} P_n^m(cos\ \theta)[\ a_{mn}cos(m\phi) + b_{mn}sin(m\phi)]$$

How many terms should be included?

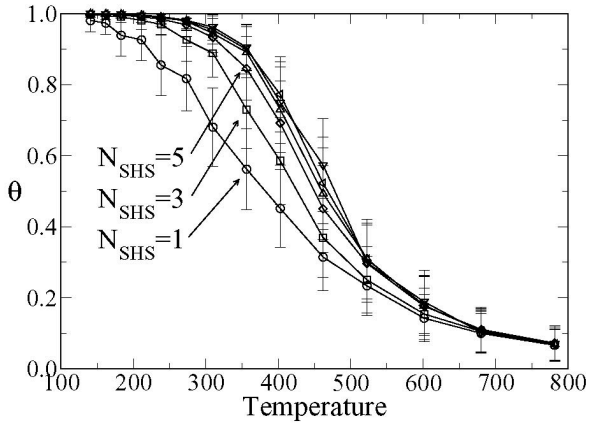



Figure 12:

- Eventually the smoothing limits and then eliminates the *cooperativity* of the structural transition!

- Van Giessen and Straub, J. Chem. Phys. (in press, 2004).

# Molecular simulations: Algorithmic and mathematical aspects

Development of coarse-grained models for protein simulations

Protein fold recognition - the protein structure prediction problem

Potentials for off-lattice Monte Carlo simulations of proteins

Application to a coil-to-helix transition in polyalanine

Vio Buchete, Alan van Giessen and Dave Thirumalai

NIH and NSF